

AGAINST BOGHOSSIAN'S CASE FOR INCOMPATIBILISM

Simon DIERIG

ABSTRACT: Two major objections have been raised to Boghossian's discrimination argument for the incompatibility of externalism and self-knowledge. Proponents of the first objection claim that thoughts about "twin water" are not relevant alternatives to thoughts about water. Advocates of the second objection argue that the ability to rule out relevant alternatives is not required for knowledge. Even though it has been shown that these two objections to Boghossian's argument are misguided, it will be argued in this essay that Boghossian's discrimination argument is nevertheless untenable. Whereas the two unsuccessful objections mentioned above each focus on one of the discrimination argument's premises in isolation, the target of my criticism of Boghossian's argument is the conjunction of its third premise and the standard incompatibilist defense of its second premise.

KEYWORDS: Paul Boghossian, discrimination argument, incompatibilism, externalism, self-knowledge

I

Boghossian's discrimination argument for the incompatibility of semantic externalism and a priori self-knowledge can be stated as follows:¹ To know a priori

¹ See Paul A. Boghossian, "Content and Self-Knowledge," *Philosophical Topics* 17, 1 (1989): 12–14; and Ted A. Warfield, "Privileged Self-Knowledge and Externalism Are Compatible," *Analysis* 52 (1992): 234–235. The term "discrimination argument" is due to Jessica Brown, *Anti-Individualism and Knowledge* (Cambridge: The MIT Press, 2004), 26. For further arguments for incompatibilism, see Akeel Bilgrami, "Can Externalism Be Reconciled with Self-Knowledge?" *Philosophical Topics* 20, 1 (1993): 240; Boghossian, "Content," 22–23; Paul A. Boghossian, "Externalism and Inference," in *Rationality in Epistemology*, ed. Enrique Villanueva (Atascadero: Ridgeview, 1992), 18–22; Paul A. Boghossian, "What the Externalist Can Know A Priori," *Proceedings of the Aristotelian Society* 97 (1997): 165–166; Jessica Brown, "The Incompatibility of Anti-Individualism and Privileged Access," *Analysis* 55 (1995): 152–155; Jessica Brown, "Reliabilism, Knowledge, and Mental Content," *Proceedings of the Aristotelian Society* 100 (2000): 118, 121, and 128; Brown, *Anti-Individualism and Knowledge*, 121 and 123; Anthony Brueckner, "Scepticism about Knowledge of Content," *Mind* 99 (1990): 448; Anthony Brueckner,

Simon Dierig

that *p* is the case, one has to be able to rule out a priori all relevant alternatives to *p*. But Oscar, our protagonist, cannot rule out a priori that he thinks that *twater* is wet. For if he were thinking that *twater* is wet, things would seem to him exactly as they seem to him in reality. (This is the standard incompatibilist justification for the second premise.²) Moreover, the proposition that Oscar thinks that *twater* is wet is a relevant alternative to the fact that he thinks that water is wet. Therefore, Oscar does not know a priori that he thinks that water is wet.³

Two major objections have been leveled at this argument. According to the first objection, the proposition that Oscar thinks that *twater* is wet is not a *relevant* alternative to the fact that he thinks that water is wet.⁴ The third premise of Boghossian's argument is therefore mistaken. Proponents of the second chief objection to the discrimination argument hold that the first premise of this argument is wrong: The ability to rule out relevant alternatives is, according to

"Knowledge of Content and Knowledge of the World," *The Philosophical Review* 103 (1994): 327–328; Keith Butler, "Externalism, Internalism, and Knowledge of Content," *Philosophy and Phenomenological Research* 57 (1997): 787–788; and Michael McKinsey, "Anti-Individualism and Privileged Access," *Analysis* 51 (1991): 15.

² See, for example, Brueckner, "Scepticism," 448.

³ Here and in what follows I assume familiarity with Putnam's and Burge's twin earth thought experiments and the relevant-alternatives approach to knowledge developed by Dretske and Goldman. The classical sources for twin earth are Hilary Putnam, "The Meaning of 'Meaning'," in *Mind, Language and Reality. Philosophical Papers, Volume 2* (Cambridge: Cambridge University Press, 1975); Tyler Burge, "Individualism and the Mental," in *Studies in Metaphysics*, eds. Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (Minneapolis: University of Minnesota Press, 1979); Tyler Burge, "Other Bodies," in *Thought and Object. Essays on Intentionality*, ed. Andrew Woodfield (Oxford: Clarendon Press, 1982); and Tyler Burge, "Intellectual Norms and Foundations of Mind," *The Journal of Philosophy* 83 (1986): 697–720. For the relevant-alternatives account, see Fred Dretske, "Epistemic Operators," *The Journal of Philosophy* 67 (1970): 1007–1023; and Alvin I. Goldman, "Discrimination and Perceptual Knowledge," *The Journal of Philosophy* 73 (1976): 771–791.

⁴ See Warfield, "Privileged Self-Knowledge," 234–235. For further discussion of this objection, see Peter Ludlow, "Externalism, Self-Knowledge, and the Prevalence of Slow Switching," *Analysis* 55 (1995): 46–49; Brown, *Anti-Individualism and Knowledge*, 138–142; Sanford Goldberg, "Brown on Self-Knowledge and Discriminability," *Pacific Philosophical Quarterly* 87 (2006): 310–311; Mikkel Gerken, "Conceptual Equivocation and Epistemic Relevance," *Dialectica* 63 (2009): 124–131; and Simon Dierig, "The Discrimination Argument Revisited," *Erkenntnis* 72 (2010): 75–78.

them, not necessary for knowledge. To support this thesis, they draw on Burge's account of self-knowledge.⁵

Although it has been argued persuasively that the two objections just outlined are not compelling, it will be shown in this essay that Boghossian's discrimination argument is nevertheless untenable.⁶ Whereas the two unsuccessful objections sketched above focus on either the first or the third premise of Boghossian's argument, that is, on one of its premises in isolation, my criticism of the discrimination argument is targeted on the conjunction of its third premise and the standard incompatibilist defense of its second premise. I will attempt to make a case for the claim that there is a conflict between the third premise—which says that the twater thought is a relevant alternative to the water thought—and the counterfactual, meant to support the second premise, that if Oscar were thinking that twater is wet, things would seem to him exactly as they seem to him in reality. Before I can present my argument for this claim, some more stage-setting is necessary. In particular, it must be explained in more detail what the notion of a *relevant* alternative amounts to and how Boghossian defends his claim that the twater thought is a *relevant* alternative to the water thought.

In his essay "Discrimination and Perceptual Knowledge," Goldman contrives the following, now famous, thought experiment:⁷ While driving in the country, Henry comes to believe that a building he drives past is a barn. Henry has normal eyesight, the building is in plain view and it is in fact a barn. Given this description

⁵ See Tyler Burge, "Individualism and Self-Knowledge," *The Journal of Philosophy* 85 (1988): 649-663; and Tyler Burge, "Our Entitlement to Self-Knowledge," *Proceedings of the Aristotelian Society* 96 (1996): 91-116. The second objection to the discrimination argument has been called "the standard strategy" of criticizing Boghossian's argument because a number of philosophers think that it is an appropriate rejoinder to this argument (see Butler, "Externalism," 780-783 and 790). Proponents of the standard strategy are, for example, Burge, Stalnaker, Falvey and Owens, and Goldberg (see Burge, "Individualism and Self-Knowledge;" Robert Stalnaker, "Narrow Content," in *Propositional Attitudes. The Role of Content in Logic, Language, and Mind*, eds. C. Anthony Anderson and Joseph Owens (Stanford: Center for the Study of Language and Information, 1990); Kevin Falvey and Joseph Owens, "Externalism, Self-Knowledge, and Skepticism," *The Philosophical Review* 103 (1994): 107-137; Sanford Goldberg, "The Dialectical Context of Boghossian's Memory Argument," *Canadian Journal of Philosophy* 35 (2005): 135-148; and Goldberg, "Brown on Self-Knowledge").

⁶ For incompatibilist responses to the first objection, see footnote 4. For a critique of the second objection, see Simon Dierig, "The Discrimination Argument and the Standard Strategy," *Grazer Philosophische Studien* 90 (2014): 213-230.

⁷ See Goldman, "Discrimination," 772-773.

of the situation, it is perfectly natural to say that Henry knows that the object he passes by is a barn. But now consider a slightly different scenario which perfectly resembles the situation just depicted with the sole exception that the countryside Henry is driving through is full of papier-mâché facsimiles of barns which cannot be distinguished under normal conditions from real barns. With regard to this new scenario, we would no longer describe Henry as knowing that the building he goes past is a barn even though it actually is a genuine barn.

According to Goldman, a relevant-alternatives epistemologist will explain why we ascribe knowledge in the first but not in the second situation as follows:⁸ In the first scenario, Henry cannot rule out the possibility that the object he drives past is a papier-mâché facsimile of a barn. But this does not prevent us from ascribing knowledge to him because, first, the proposition that the building he is looking at is a papier-mâché barn is not a *relevant* alternative to the fact that the object in question is a genuine barn and, second, it is not every conceivable alternative, but only *relevant* alternatives, that must be excluded in order for a knowledge ascription to be true.

The second situation differs importantly from the first in that the proposition that the object Henry goes past is a papier-mâché facsimile of a barn is now a relevant alternative to the fact that the building before him is a barn. What makes this proposition a *relevant* alternative is the presence of papier-mâché facsimiles of barns in Henry's surroundings. Yet, if the proposition that the object Henry passes by is a papier-mâché barn is a relevant alternative, one has to conclude from his inability to rule out this proposition and the principle that knowledge requires the ability to rule out relevant alternatives that, in the second situation, he does not know that the building he is looking at is a barn.

To apply the relevant-alternatives approach to the twin earth scenario and to support the third premise of his argument, Boghossian imagines that our protagonist Oscar travels back and forth between earth and twin earth and stays on twin earth long enough to acquire the twin earthian concept *twater*.⁹ On earth Oscar thinks that water is wet, whereas on twin earth he thinks that *twater* is wet. Just as the *actual presence* of papier-mâché facsimiles of barns in the area Henry is driving through makes the proposition that the object he is looking at is a papier-mâché barn a *relevant* alternative, so the fact that Oscar *actually* thinks on twin

⁸ See Goldman, "Discrimination," 774–775.

⁹ See Boghossian, "Content," 13–14.

earth that twater is wet makes the proposition that he thinks that twater is wet a *relevant* alternative to the fact that he thinks on earth that water is wet.

II

Having explained the notion of a relevant alternative and, in particular, the relation between relevance and “actuality,” I can now raise my objection to the discrimination argument. It consists of four steps. First, if the proposition that Oscar thinks that twater is wet, rather than that water is wet, is a *relevant* alternative, it must be “actual,” that is, Oscar must have thought, at some time in his not-too-remote past, that twater is wet, but not that water is wet. From the consequent of this conditional it follows that our protagonist has recently been on a planet on which there is no H₂O, but only XYZ. For want of better terminology, I shall from now on refer to this planet as “twin earth.” To the *counterfactual* counterpart of actual earth which *only* differs from earth in that all H₂O is replaced with XYZ I shall from now on refer as “counterfactual twin earth.” So armed, the claim which corresponds to the first step of my objection can be stated in the following way: If the proposition that Oscar thinks the twater instead of the water thought is relevant, then it is actual, that is, he has recently been on twin earth, thinking that twater is wet.

Second, if twin earth does not exist in reality, the closest counterfactual situation in which our protagonist does not think the water but rather the twater thought is one in which he lives on counterfactual twin earth, i.e., on a counterfactual counterpart of actual earth which only differs from earth in that all H₂O is replaced with XYZ. But if twin earth exists in reality, it can be argued that the closest counterfactual situation in which Oscar thinks the twater instead of the water thought is a situation in which he lives on twin earth, rather than on counterfactual twin earth. The argument runs like this: If twin earth exists in reality, the closest counterfactual situation in which Oscar lives on twin earth is closer to the actual world than the closest counterfactual situation in which huge amounts of a certain substance, viz. H₂O, are replaced on earth with a different substance, viz. XYZ, and Oscar lives on earth. But the set of counterfactual situations in which our protagonist thinks the twater instead of the water thought is identical to the set of counterfactual situations of the two kinds just mentioned. Thus, if twin earth exists in reality, the closest counterfactual situation in which

Simon Dierig

Oscar has the twater instead of the water thought is one in which he lives on twin earth, rather than on the envisaged counterpart of earth.

Third, on the assumption that twin earth exists in the actual world, both earth and twin earth exist in reality. But, for all we know, there are no two planets in the actual world which are phenomenal duplicates of each other. Hence, provided that twin earth exists in reality, twin earth is (unlike counterfactual twin earth) not a phenomenal duplicate of earth. From this it follows that (a) if twin earth is real, the following counterfactual is true: If Oscar were on twin earth, he would not have the same “pure phenomenological feels”¹⁰ as he actually has. But if he would not have the qualitative mental states he actually has if he were on twin earth, then it is a fortiori true that he would not have the qualitative mental states he actually has if he were not only on a different planet than in reality (i.e., on twin earth rather than on earth) but, moreover, had different thoughts than in reality (i.e., the twater instead of the water thought). The following conditional is therefore true as well: (b) If the counterfactual which makes up the consequent of the conditional (a) is true, the subsequent counterfactual is also true: If Oscar were on twin earth, thinking the twater instead of the water thought, he would not have the qualitative mental states he actually has. The conditionals (a) and (b) together imply (c) that given that twin earth exists in the actual world, the following counterfactual is true: If our protagonist were on twin earth, thinking the twater instead of the water thought, he would not have the qualitative mental states he actually has.

Fourth, from what was said in the second step of my objection it can be inferred that if twin earth exists in reality, Oscar would be on twin earth if he had the twater instead of the water thought. From this conditional and the conditional argued for in step 3 one can conclude, using only propositional logic and the inference rule

$$\chi \Box \rightarrow \varphi, \chi \& \varphi \Box \rightarrow \psi \Rightarrow \chi \Box \rightarrow \psi^{11}$$

that given that twin earth exists in the actual world, Oscar would not have the qualitative mental states he actually has if he had the twater instead of the water thought. Finally, from this claim and the contention argued for in the first step of my objection it follows that if the proposition that our protagonist thinks the twater instead of the water thought is relevant, it is wrong that he would have the

¹⁰ Burge, “Individualism and Self-Knowledge,” 653.

¹¹ See David Lewis, *Counterfactuals* (Oxford: Basil Blackwell, 1973), 32–35.

qualitative mental states he actually has if he had the twater instead of the water thought. Thus, either this counterfactual is mistaken, or thinking the twater instead of the water thought is not a relevant alternative to thinking the water thought. That is, one of two has to go: either the standard incompatibilist justification for the second premise of the discrimination argument or its third premise.

III

The objection to the discrimination argument raised in the previous section will now be elaborated and further clarified by responding to a number of incompatibilist rejoinders. *First rejoinder:* One premise of the objection presented above is that the set of counterfactual situations in which Oscar thinks that twater is wet, rather than that water is wet, comprises only situations in which he lives on twin earth and situations in which huge amounts of a certain substance, viz. H₂O, have been replaced on earth with a different substance, viz. XYZ, and he lives on earth. But there is a third category of counterfactual situations in which our protagonist has the twater instead of the water thought, namely situations in which he is on his journey from twin earth to earth.

Let us grant for the sake of the argument that if twin earth exists in reality, the closest "journey situation" is closer to the actual world than all other counterfactual situations of the three categories in question. It follows that on the assumption that twin earth is actual, the closest counterfactual situation in which Oscar has the twater instead of the water thought is one in which he is on his journey from twin earth to earth. But on this journey our protagonist would not have the same "pure phenomenological feels" as he has in reality because he would be exposed to different sensory input. Therefore, given that twin earth exists in reality, Oscar would have different qualitative mental states if he had the twater instead of the water thought.

Second rejoinder: An advocate of the objection leveled in the preceding section must assume that XYZ is not water. For if XYZ were water, Oscar would not only have the twater but also the water belief on twin earth (and of course also on counterfactual twin earth). Yet it might be argued that XYZ *is* water because it is a colourless, tasteless etc. liquid.

I reply that if the English word "water" is synonymous with the expression "colourless, tasteless etc. liquid," the same will be true for the Twin English word "water." From this it follows that the English as well as the Twin English

expression “water” have the same meaning. Since the English neologism “twater” is stipulated to have the same meaning as the Twin English word “water,” it can be concluded that the English expressions “water” and “twater” are synonymous. But synonymous expressions can be substituted *salva veritate* in belief contexts. Thus, it is impossible that Oscar believes that twater is wet without at the same time believing that water is wet. It follows that the counterfactual “If Oscar had the twater instead of the water thought, he would have the qualitative mental states he actually has” lacks a truth value and cannot therefore be employed to justify the second premise of the discrimination argument.

One might object that the principle that synonymous expressions can be substituted *salva veritate* in belief contexts must be dismissed because Mates has shown that it is faulty regarding higher-order belief contexts and, more importantly, Burge has argued that it is even wrong regarding simple, first-order belief contexts.¹² According to Burge, a person who misunderstands arthritis to be simply a rheumatoid ailment can believe that she has arthritis in her thigh without believing that she has an inflammation of joints in her thigh. In response to this objection, it suffices to point out that, first, the belief contexts in our example are, unlike the belief contexts in Mates’ examples, not higher-order contexts and that, second, our example does not involve incomplete understanding of one of the pertinent expressions, as does Burge’s arthritis case.

Third rejoinder: The first step of the objection to Boghossian’s discrimination argument is flawed. It does not follow from the claim that Oscar has thought, at some time in his not-too-remote past, the twater instead of the water thought that he has recently been on twin earth before travelling to earth. Saying the sentence “Water is wet” inwardly to oneself while having causal contact to twater on twin earth is not the only way of thinking the twater instead of the water thought. Another way is to say the sentence “*Twater* is wet” inwardly to oneself, regardless of whether one has causal contact to twater.

The neologism “twater” is only known to philosophers who are acquainted with twin earth thought experiments of the Putnam–Burge variety. The third rejoinder therefore works only for those few cases in which our protagonist is such a philosopher. But this means that my objection to the discrimination argument is

¹² See Benson Mates, “Synonymity,” in *Semantics and the Philosophy of Language*, ed. Leonard Linsky (Urbana: University of Illinois Press, 1952), 125; Tyler Burge, “Belief and Synonymy,” *The Journal of Philosophy* 75 (1978): 119–138; and Burge, “Individualism and the Mental.”

successful in the vast majority of cases in which no knowledge of the Putnam–Burge story about twin earth is available.

Fourth rejoinder: Boghossian argues that if the proposition that his protagonist thinks the twater instead of the water thought is actual, then it is relevant. What is claimed to be true in the first step of the objection raised above is, however, the *converse* conditional that if the proposition just mentioned is relevant, it has to be actual. Whereas the former conditional, endorsed by Boghossian, is fairly unproblematic (because it can be supported by analogy to the barn example), the latter conditional is much more dubious. For if the proposition in question is relevant, one can explain this by drawing on the fact that Oscar *believes* this proposition to be true. It is therefore wrong to suppose that if the proposition in question is relevant, this can only be explained with recourse to the assumption that it is actual. Thus, we lack any reason for claiming that the proposition in question has to be actual if it is relevant.

One way to impugn this rejoinder is to deny Lewis' "rule of belief," according to which a proposition is relevant if the protagonist believes that it is true.¹³ But even if Lewis' rule is correct, it can be shown as follows that the fourth rejoinder fails. Remember that compatibilism is the view that both externalism and the doctrine that we have a priori self-knowledge are true. In cases in which the protagonist does not believe that she has a particular thought this view is obviously wrong since in these cases she does of course not possess a priori knowledge of the thought in question. Compatibilism and its negation, incompatibilism, should therefore be construed as claims which concern only situations in which the protagonist believes that she has the thought in question. With regard to our protagonist Oscar this means that he must be envisaged as believing that he thinks that water is wet. However, from the assumption that he has this belief it can be inferred that he does not believe that he does not think that water is wet. From this it follows, in turn, that he does not believe the proposition "Oscar does not think that water is wet but rather that twater is wet." If this proposition is relevant, this cannot therefore be explained by falling back on the assumption that our protagonist believes it to be true. In brief, the fourth incompatibilist rejoinder founders.

¹³ See David Lewis, "Elusive Knowledge," in *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press, 1999), 428–429.

IV

Fifth rejoinder: Admittedly, if the proposition in question is relevant, this cannot be the case because our protagonist believes it to be true. But there are plenty of other criteria of relevance apart from Goldman's principle of actuality and Lewis' rule of belief. As long as it has not been shown that one cannot explain by invoking one of them why the proposition in question is relevant, if it is relevant, one cannot legitimately reason from the claim that this proposition is relevant to the contention that it is actual.

To counter this challenge, I will examine in this section those principles of relevance advanced by the chief proponents of the relevant-alternatives account of knowledge which have not yet been considered. It will be argued that these principles are (with one exception) either untenable because they invite scepticism or contextualism, or cannot be used to explain why the proposition in question is relevant—if it is relevant. It follows that in case the proposition that Oscar thinks the twater instead of the water thought is relevant, this must be explained with recourse to Goldman's principle of actuality, the only remaining principle of relevance, and to the assumption that it is fulfilled in the case of our protagonist (I am simplifying somewhat). The upshot of my argument is that if the proposition in question is relevant, it must be actual. In other words, the claim which corresponds to the first step of my objection to the discrimination argument is true.

Let me begin my inquiry into the not yet discussed rules of relevance with the principle of relevancy put forward by Dretske in his seminal essay "Epistemic Operators." Dretske writes: "A relevant alternative is an alternative that might have been realized in the existing circumstances if the actual state of affairs had not materialized."¹⁴ There are (at least) two ways to understand this statement. On the face of it, Dretske claims that an alternative q to p is relevant iff were p wrong, q *might* be the case. If this claim is true, sceptical hypotheses like Descartes' deceiving-god or Putnam's brain-in-a-vat hypothesis are relevant alternatives. To see this, suppose that p is the true proposition that I am sitting on a chair in Jerusalem and q is the proposition that I am a brain in a vat. It follows from Dretske's relevance criterion, as interpreted above, that the latter proposition is relevant iff the following is true: If I were not sitting on a chair in Jerusalem, I *might* be a brain in a vat. Intuitively, this counterfactual is true. Thus, the brain-

¹⁴ Dretske, "Epistemic Operators," 1021.

in-a-vat hypothesis is a relevant alternative. But this is, of course, capitulation to scepticism.

The way Dretske applies his criterion of relevance to examples, however, suggests the following second reading of it: An alternative q to p is relevant iff were p wrong, q *would* be the case.¹⁵ Interpreted in this way, Dretske's principle does not invite scepticism. For the counterfactual "If I were not sitting on a chair in Jerusalem, I *would* be a brain in a vat" is clearly wrong. But can one draw on Dretske's principle to explain why the proposition that Oscar thinks the twater instead of the water thought is relevant—if it is relevant? This question must be answered in the negative. For the closest counterfactual situation in which our protagonist does not think that water is wet is not a situation in which he is on twin earth or on counterfactual twin earth and therefore thinks the twater instead of the water thought, but is rather a situation in which he lives on earth and says a different sentence inwardly to himself. Accordingly, the counterfactual "If Oscar did not think that water is wet, he would think the twater instead of the water thought" is wrong.

In sum, the first version of Dretske's principle of relevancy is mistaken since it leads to scepticism. The second version of Dretske's principle does not invite scepticism but, since the counterfactual in question is wrong, cannot be used to explain why the proposition "Oscar thinks the twater instead of the water thought" is relevant, if it is relevant.

One may level the objection that the "might" in Dretske's principle of relevance has been misunderstood. Following Lewis, one may claim that the counterfactual "If p were wrong, q *might* be the case" has to be analyzed as "*It is false that* if p were wrong, q *would not* be the case."¹⁶ Given this analysis of the might-counterfactual (in terms of the would-counterfactual), the conditional "If I were not sitting on a chair in Jerusalem, I might be a brain in a vat" is not true, as alleged above, but false. For the counterfactual "If I were not sitting on a chair in Jerusalem, I would not be a brain in a vat" is true. But in case the conditional "If I were not sitting on a chair in Jerusalem, I might be a brain in a vat" is wrong, the first variant of Dretske's principle does not imply, together with true additional premises, that the brain-in-a-vat hypothesis is relevant. Hence, the first version of Dretske's criterion, correctly understood, does not lead to scepticism.

¹⁵ See Dretske, "Epistemic Operators," 1021, fn. 6; see also Robert Nozick, *Philosophical Explanations* (Oxford: Clarendon Press, 1981), 175.

¹⁶ See Lewis, *Counterfactuals*, 2.

But can one fall back on this criterion to explain why the proposition that Oscar has the twater instead of the water thought is relevant—if it is relevant? Intuitively, it is clear that Oscar would not have the twater instead of the water thought if he did not think that water is wet. Given Lewis' analysis, it follows that it is false that if our protagonist did not think that water is wet, he might think the twater instead of the water thought. Thus, in case Dretske's counterfactual principle is understood in Lewis' style, one cannot use it to explain why the proposition "Oscar thinks the twater instead of the water thought" is relevant, if it is relevant.

In a later essay, Dretske advances a modal criterion for relevance, according to which an alternative p is relevant iff it is a genuine possibility that p is the case.¹⁷ Dretske does not explain in more detail what he has in mind when he talks of "genuine possibilities." All he says is that possibility in his sense is "objective" and does not amount to logical possibility. From this it follows that it can be understood either as metaphysical or as nomological possibility. If it is understood as metaphysical possibility, it can be inferred from Dretske's principle, together with the true claim that the proposition "I am deceived by an evil demon or a mad scientist" is metaphysically possible, that this proposition is a relevant alternative. That is, if the notion of possibility incorporated in Dretske's principle is the notion of metaphysical possibility, this principle invites scepticism and is therefore untenable.

If, on the other hand, possibility in Dretske's sense is nomological possibility, his principle cannot be invoked to explain why the proposition "Oscar thinks the twater instead of the water thought" is relevant—if it is relevant. This can be shown as follows: Thinking the twater instead of the water thought requires causal contact to XYZ, that is, to a substance which has the same phenomenological properties as H₂O, but a completely different chemical composition. But, as far as we can tell, it contradicts the laws of nature that such a substance exists. Thus, it is nomologically impossible that our protagonist thinks the twater instead of the water thought.¹⁸

¹⁷ See Fred Dretske, "The Pragmatic Dimension of Knowledge," *Philosophical Studies* 40 (1981): 376–378.

¹⁸ In addition to the modal criterion of relevance just discussed, Dretske examines, in the paper mentioned above, four other rules of relevance (see Dretske, "The Pragmatic Dimension," 373–376). None of them can be employed to explain why the proposition "Oscar thinks the twater instead of the water thought" is relevant, if it is relevant. Due to limitations of space, I cannot

A counterfactual principle of relevance different from the one proposed by Dretske has been suggested by Luper. He holds that “an alternative to p , A , is *relevant* (relative to S and S 's situation) if and only if: RA: In S 's circumstances, A might hold (i.e., it is false that: given S 's circumstances A would not hold).”¹⁹ This principle may or may not be true. But even if true, one cannot explain on its basis why the proposition (or alternative) that Oscar has the twater instead of the water thought is relevant—if it is relevant. To see this, it has to be borne in mind that Oscar's circumstances encompass his interactions with his H₂O-containing environment. From this it follows that given our protagonist's circumstances he would not have the twater instead of the water thought. In other words, the right-hand side of Luper's principle is wrong if A is the proposition “Oscar has the twater instead of the water thought.” One cannot therefore explain with the aid of Luper's principle why this proposition is relevant, if it is relevant.

In addition to his principle of actuality, Goldman espouses three other rules of relevance.²⁰ The first of them says that if it is likely or probable that a particular alternative obtains (rather than the actual state of affairs), it amounts to a relevant alternative. Can one explain by drawing on this rule why the proposition “Oscar thinks the twater instead of the water thought” is relevant—if it is relevant? The notion of probability presupposed here is not the notion of objective probability—i.e., relative frequency—but rather the notion of subjective probability—i.e., degree of belief. For our protagonist's thinking the twater instead of the water thought at a particular time is a particular event (or proposition), and particular events cannot be objectively probable, but only subjectively probable. Only repeatable event or proposition types can be objectively probable.

The notion of subjective probability, or degree of belief, can in our context be understood in two ways, depending on whether the pertinent degree of belief is that of our thought experiment's protagonist or that of us philosophers who think about the thought experiment. In the following, it will be argued that the proposition in question is not subjectively probable, no matter which of the two interpretations of subjective probability is chosen.

elaborate on this here.

¹⁹ Steven Luper, “Dretske on Knowledge Closure,” *Australasian Journal of Philosophy* 84 (2006): 380; see also Steven Luper(-Foy), “The Epistemic Predicament: Knowledge, Nozickian Tracking, and Scepticism,” *Australasian Journal of Philosophy* 62 (1984): 46–48.

²⁰ See Goldman, “Discrimination,” 776.

Suppose, first, that the relevant degree of belief is our protagonist's degree of belief. In my reply to the fourth rejoinder, I argued that Oscar must be envisaged as believing that he thinks that water is wet. But if he has this belief, he does not believe that he does not think that water is wet. From this it can be inferred that Oscar does not believe that the proposition "Oscar does not think that water is wet, but rather that twater is wet" is true. His degree of belief in this proposition is therefore not (sufficiently) high. Given the first interpretation of subjective probability, the proposition in question is accordingly not subjectively probable.

Suppose, then, that the second interpretation of subjective probability is true: The pertinent degree of belief is the degree of belief of us philosophers who think about the thought experiment. From the fact that we know for sure that Oscar thinks the water thought it can be concluded that we also know for sure that the proposition "Oscar has the twater instead of the water thought" is wrong. Our degree of belief in this proposition is therefore zero. Not only on the first, but also on the second interpretation of subjective probability this proposition is thus not subjectively probable. But this means that one cannot use Goldman's rule of probability to explain why the proposition in question is relevant, if it is relevant.

According to Goldman's second principle of relevance, a proposition is relevant if the situation in which it obtains is similar to the actual situation. There are two problems with this principle. The first problem is that there is not only one possible situation in which a given proposition obtains. Therefore, the question arises which possible situation or set of possible situations is meant. One natural answer would be that the possible situations to be specified are those which are closest to the actual world. The second principle would then read as follows: A proposition is relevant if the closest possible worlds in which it obtains are similar to the actual world.

There is, however, a second difficulty with Goldman's principle which cannot be resolved that easily. It pertains to the concept of *similarity* between counterfactual situations or possible worlds. We have a clear idea of when a situation or possible world resembles another *in a certain respect*. We have a less clear, but still fairly clear idea of when a possible world is *more similar* to the actual world than another possible world. But we are almost always at a loss when it comes to deciding whether a particular world is *similar simpliciter* to another possible world or to reality. To illustrate this difficulty, consider, once again, counterfactual twin earth. This situation is blatantly more similar to reality than, for example, the deceiving-god or brain-in-a-vat scenario. Moreover, it resembles

reality in a number of respects, such as the phenomenal properties of the stuff called "water" by Oscar's compatriots, but differs from reality in a number of other respects, such as the chemical composition of the stuff called "water" in our protagonist's language community. All these "similarity claims" seem fairly unproblematic. But if forced to tell whether counterfactual twin earth is similar *tout court* to the actual world, we are at a loss. We can compare counterfactual twin earth with other situations regarding their similarity to reality; and we can say in which respects it resembles, or differs from, actual earth. But when it comes to the all-or-nothing question whether counterfactual twin earth is similar to reality, we cannot come up with an answer.

It might be objected that similarity is mostly understood as coincidence in some (but not necessarily all) properties and that counterfactual twin earth is therefore clearly similar to reality. To counter this objection, it suffices to point out that, according to the proposed explanation of similarity, even brain-in-a-vat scenarios are similar to the actual world. Given Goldman's principle of similarity, it follows that the proposition that I am a brain in a vat is a relevant alternative. Thus, if the proposed explanation of similarity is correct, Goldman's principle of similarity invites scepticism and must therefore be rejected.

Goldman's third principle of relevance says that a proposition which is taken seriously by the ascriber of knowledge is relevant. This principle presupposes contextualism regarding knowledge, that is, the view that the truth conditions of knowledge attributions depend on the linguistic and psychological context of the knowledge ascriber. Powerful objections have been raised to this doctrine.²¹ Whoever finds them convincing cannot endorse Goldman's third principle of relevance.

A more elaborate version of the probabilistic criterion of relevance discussed above has been defended by Cohen. It can be formulated as follows: A particular alternative or proposition *p* is relevant if the probability of *p* conditional on the subject's evidence and certain features of the circumstances is sufficiently high.²² For the reasons already explained, the notion of probability which occurs in this principle must be understood epistemically. Cohen's principle of relevance can

²¹ See, for example, John Hawthorne, *Knowledge and Lotteries* (Oxford and New York: Oxford University Press, 2004); and Jason Stanley, *Knowledge and Practical Interests* (Oxford and New York: Oxford University Press, 2005).

²² See Stewart Cohen, "How to Be a Fallibilist," in *Epistemology*, ed. James E. Tomberlin (Atascadero: Ridgeview, 1988), 95 and 102.

therefore be restated in the following way: A proposition p is relevant if the subject's evidence and certain features of the circumstances confirm p to a sufficiently high degree.

In the second variant of the barn example, the protagonist's evidence does not confirm the proposition that the object he is driving by is a fake barn. For he has no clue that he is driving through a countryside full of papier-mâché facsimiles of barns. If Cohen's criterion is nonetheless fulfilled regarding the proposition in question, this must be because certain features of Henry's circumstances confirm this proposition. But they can do this only if they are constituted by a number of objects which are fake barns. In brief, if Cohen's principle is fulfilled regarding the proposition in question, the property of being a fake barn must be "actual."

In the same vein, it can be argued that if Cohen's criterion is fulfilled regarding the proposition "Oscar thinks the twater instead of the water thought," this proposition is actual: Oscar's evidence does not confirm this proposition since he does not know anything about twin earth, twater etc. If Cohen's principle is nonetheless fulfilled regarding the proposition in question, this must be because certain features of our protagonist's circumstances confirm this proposition. But they can do this only if they are constituted by Oscar's having thought the twater instead of the water thought many times in his recent past. Thus, if Cohen's principle is fulfilled regarding the proposition in question, this proposition is actual.

In addition to his "external" probabilistic principle of relevance, Cohen puts forward the following "internal" principle of relevancy: "an alternative (to q) h is relevant, if S lacks sufficient evidence (reason) to deny h , i.e., to believe not- h ."²³ Because this principle invites scepticism, Cohen modifies it as follows: An alternative h is relevant if, first, S lacks sufficient evidence to believe not- h and, second, it is not the case that not- h is intrinsically rational, where a proposition is intrinsically rational iff it can be rational to believe this proposition without possessing evidence for it.²⁴

Is Cohen's modified principle doing any better than his original principle when it comes to the issue of scepticism? At first glance, it might seem so. According to Cohen, it is rational to believe that we are not brains in a vat even though we have no evidence for this belief.²⁵ The hypothesis that we are not brains

²³ Cohen, "How to Be a Fallibilist," 103, see also 102.

²⁴ See Cohen, "How to Be a Fallibilist," 111-113.

²⁵ See Cohen, "How to Be a Fallibilist," 112.

in a vat is therefore intrinsically rational. It follows that the second conjunct of the amended principle's antecedent is wrong (if the brain-in-a-vat hypothesis is substituted for "h"). Hence, one cannot draw on Cohen's modified principle to argue that the brain-in-a-vat hypothesis is relevant.

The crucial premise underlying this line of argument is the claim that we do not have evidence for the belief that we are not brains in a vat or the victims of a deceiving god. But this claim is debatable. Descartes' proof of a benevolent god is meant to show that it is impossible that god deceives us; and Putnam invokes externalism to argue that it is impossible that we are brains in a vat. But even if we are reluctant to rely on Cartesian theology or on an externalist semantics, the claim that we do not possess evidence which counts against sceptical hypotheses of the brain-in-a-vat variety is questionable. Consider the following reasoning: I am sitting on a chair in Jerusalem. If I am sitting on a chair in Jerusalem, I am not a brain in a vat. Therefore, I am not a brain in a vat. This reasoning may not show that I *know* that I am not a brain in a vat. It may not constitute a *proof* of the contention that the brain-in-a-vat hypothesis is wrong. But it clearly provides me with a *reason*, albeit not a conclusive one, for believing that I am not a brain in a vat. The two premises of the above reasoning constitute *evidence*, even though not conclusive evidence, for the contention that the brain-in-a-vat hypothesis is false. Cohen's argument for the claim that the hypothesis that we are not brains in a vat is intrinsically rational is therefore unconvincing. Since no other argument for this claim is in the offing and the burden of proof is on those who endorse it, one can legitimately conclude that it is mistaken. In short, not only the original but also the modified version of Cohen's internal principle leads to scepticism.²⁶

²⁶ The rules of relevance advanced by Lewis include necessary as well as sufficient conditions for relevance (see Lewis, "Elusive Knowledge," 426–435). The former are not pertinent in our context. The latter comprise the "rule of actuality," the "rule of belief," the "rule of resemblance" and the "rule of attention." It has already been shown that the rule of belief cannot be used to explain why the proposition that Oscar thinks the twater instead of the water thought is relevant, if it is relevant (see my reply to the fourth rejoinder). As to the rules of actuality and of resemblance, they only make sense in Lewis' ontological framework, in which alternatives are not construed as propositions, but rather as possibilities. Finally, the rule of attention presupposes contextualism and is therefore problematic. The principle of relevance proposed by Stine is a necessary condition for relevance and is therefore not pertinent here (see G. C. Stine, "Skepticism, Relevant Alternatives, and Deductive Closure," *Philosophical Studies* 29 (1976): 252–253).

Simon Dierig

To sum up, the principles of relevance put forward by advocates of the relevant-alternatives approach to knowledge—with the exception of Goldman’s principle of actuality and Cohen’s principle of probability—are either untenable because they invite scepticism or contextualism, or cannot be used to explain why the proposition that our protagonist thinks the twater instead of the water thought is relevant—if it is relevant. It follows that if this proposition is relevant, this must be explained either with recourse to Goldman’s principle of actuality and the assumption that it is fulfilled regarding the proposition in question, or with recourse to Cohen’s principle of probability and the contention that *it* is fulfilled with regard to this proposition. But if Cohen’s principle is fulfilled regarding the proposition in question, the same is true for Goldman’s principle (as has been shown above). Thus, if the proposition that our protagonist thinks the twater instead of the water thought is relevant, it is actual. In other words, the claim which corresponds to the first step of my objection to the discrimination argument is true.

V

Sixth rejoinder: It must be admitted that the discrimination argument, as it has been reconstructed at the beginning of this essay, is untenable. But Boghossian’s argument can be saved if one reformulates it by using the idiom of mental events and their contents. Consider the following *content-based* variant of the discrimination argument (*e* is meant to be the mental event which takes place “in” our protagonist when he occurrently thinks that water is wet):

(P₁) To know a priori that *p* is the case, one must be able to rule out a priori all relevant alternatives to *p*.

(P₂) Oscar cannot rule out a priori that *e* has the content “Twater is wet.”

(P₃) The proposition that *e* has the content “Twater is wet” is a relevant alternative to the fact that *e* has the content “Water is wet.”

(C₁) So Oscar does not know a priori that *e* has the content “Water is wet.”

(P₄) Oscar knows a priori *that e has the content “Water is wet” if he thinks that water is wet.*

(P₅) Closure: If, first, one knows a priori that *p* and, second, one knows a priori *that if p, then q*, then one knows a priori that *q*.

(C₂) So Oscar does not know a priori that he thinks that water is wet.

A proponent of this version of the discrimination argument is confronted with the following dilemma. Either it is possible that *e* has both the content "Water is wet" and the content "Twater is wet," then the premise (P₃) is mistaken; or it is impossible that *e* has both contents just mentioned, then the standard incompatibilist defense of the premise (P₂) is not true.

The second horn of this dilemma can be substantiated as follows: Mental events have their contents necessarily. It is impossible that they exist without having the content they actually have. The proposition "e has the content 'Twater is wet'" therefore implies the proposition "e has the content 'Water is wet' as well as the content 'Twater is wet'." However, by assumption it is impossible that *e* has these two contents simultaneously. Thus, it is impossible that *e* has the content "Twater is wet." But this means that the antecedent of the counterfactual "If *e* had the content 'Twater is wet,' everything would seem to Oscar as it seems to him in reality" is impossible. It follows that the truth value of this counterfactual is indeterminate and that it cannot therefore be used to support the premise (P₂). In sum, the discrimination argument cannot be saved by employing content terminology.

Seventh rejoinder: The discrimination argument, as it has been understood so far, makes use of the notion of *ruling out* a priori certain alternatives, which are construed as propositions of a certain kind. Yet the basic intuition behind Boghossian's argument is the observation that one cannot *distinguish* a priori the water thought from the twater thought. The discrimination argument should therefore be construed as making use of the notion of an ability to *distinguish* a priori between thoughts, rather than of an ability to *rule out* a priori certain propositions. In more detail, Boghossian's argument should be reconstructed as follows: To know a priori that one thinks the thought *t*, one has to be able to distinguish *t* a priori from all thoughts *t'* such that the proposition that one thinks *t'* rather than *t* is a relevant alternative to the fact that one thinks *t*. But our protagonist Oscar cannot distinguish a priori his thought that water is wet from the thought which he would have instead of the water thought if he were on counterfactual twin earth, thinking that twater is wet. For if he were on counterfactual twin earth, thinking that twater is wet, the thought which he would think instead of the water thought, call it "*t**," would have the same "pure phenomenological feels" as his real-world thought that water is wet. Moreover, the proposition that Oscar thinks *t** rather than the water thought is a relevant

alternative to the fact that he thinks the water thought. Therefore, our protagonist does not know a priori that he thinks that water is wet.

The questionable premise of this line of argument is the third one. The proposition that Oscar thinks t^* rather than the water thought is only relevant if it is actual, that is, if he has thought t^* rather than the water thought in the recent past.²⁷ But the opponent of the discrimination argument can deny that Oscar has thought t^* in the recent past. To dispute this, he need not deny that our protagonist has recently thought that *twater* is wet. On the contrary, he may grant that Oscar has thought that *twater* is wet in the not-too-remote past. Yet the compatibilist can insist that this thought is not identical to t^* , that is, to the thought which Oscar would have instead of the water thought if he were on counterfactual twin earth, thinking that *twater* is wet. Of course not every thought with the content that *twater* is wet is identical to t^* . It is the incompatibilist who needs to show that one of our protagonist's past thoughts with the content that *twater* is wet is identical to t^* . As long as the incompatibilist does not succeed in doing this, the compatibilist is justified in claiming that the proposition that Oscar thinks t^* rather than the water thought is not actual and therefore not relevant.

Eighth rejoinder: It has to be conceded that one cannot devise a convincing version of the discrimination argument by using the notion of an ability to distinguish one thought from another. But why not contrive a variant of Boghossian's argument which is based on the notion of an ability to distinguish *situations*, rather than *thoughts*? Consider the following line of reasoning:²⁸ To know a priori that one thinks that *p*, one has to be able to distinguish a priori the actual situation from all relevant counterfactual situations in which one does not think that *p*. But our protagonist Oscar cannot distinguish a priori the actual situation from the counterfactual situation in which he lives on counterfactual twin earth. For the "pure phenomenological feels" he has in the actual situation and those he has in this counterfactual situation are the same. Moreover, the counterfactual situation in question is a relevant counterfactual situation in which our protagonist does not think the water but rather the *twater* thought. Therefore, Oscar does not know a priori that he thinks that water is wet.

What is problematic about this argument is the claim that the counterfactual situation in which our protagonist lives on counterfactual twin earth is relevant. Why should one think that this counterfactual situation is relevant? One answer

²⁷ This claim can be established along the lines put forward in section IV.

²⁸ See Brown, *Anti-Individualism and Knowledge*, 37-45.

would be: because it is similar to reality. But does counterfactual twin earth resemble the actual world? As it has been pointed out in the preceding section, we can compare counterfactual twin earth with other counterfactual situations regarding their similarity to reality; and we can say in which respects it resembles, or differs from, actual earth. But when it comes to the question whether counterfactual twin earth is similar *tout court* to reality, we are at a loss. One cannot therefore argue that counterfactual twin earth is relevant because it is similar to the actual world.

A second answer to the above question would be that counterfactual twin earth is relevant because it is metaphysically possible. But this answer will not do either. Metaphysical possibility does not imply relevance because otherwise the brain-in-a-vat scenario, being metaphysically possible, would be relevant as well, which would be capitulation to scepticism. As to nomological possibility, it may imply relevance, but this does not help the proponent of the discrimination argument for counterfactual twin earth is not nomologically possible.

A third answer to the question raised above would be that if the proposition that our protagonist thinks the twater instead of the water thought is actual, counterfactual situations in which this proposition is true are relevant. But does one really want to claim that *all* counterfactual situations of the kind just specified are relevant if the proposition in question is actual? This question has to be answered in the negative for an affirmative answer would invite scepticism. Therefore, the question arises which counterfactual situations of the kind specified are relevant if the proposition in question is actual. The only nonarbitrary answer seems to be: those counterfactual situations of the kind specified *which are closest to reality*. One therefore arrives at the following rule of relevance: If the proposition that Oscar thinks the twater instead of the water thought is actual, the *closest* counterfactual situations in which this proposition is true are relevant. Can one fall back on this version of the principle of actuality to make a case for the claim that counterfactual twin earth is relevant?

My argument for a negative answer to this question consists of two steps. In the first step, it is shown that if the proposition that Oscar thinks the twater instead of the water thought is actual, counterfactual twin earth is not among the closest counterfactual situations in which this proposition is true. My argument for this claim rests upon three premises: First, if the proposition in question is actual, it is relevant. Second, as has been argued in section II, if the proposition in question is relevant, the closest counterfactual situations in which it is true are situations in

Simon Dierig

which our protagonist does not have the same “pure phenomenological feels” as in reality. Third, the qualitative mental states Oscar has on counterfactual twin earth and those he has in reality are the same. From these three premises it follows that if the proposition in question is actual, counterfactual twin earth is not among the closest counterfactual situations in which this proposition is true.

That brings me to the second step of my argument. The above version of the principle of actuality can only be employed to show that counterfactual twin earth is relevant if, first, the proposition that Oscar thinks the twater instead of the water thought is actual and, second, counterfactual twin earth is among the closest counterfactual situations in which this proposition is true. But it has been demonstrated in the first step of my argument that at least one of these two claims is wrong. Thus, one cannot invoke the principle of actuality to argue that counterfactual twin earth is relevant.

VI

Let me conclude by summarizing briefly what I have attempted to show in this essay. In section II, it has been argued that Boghossian’s discrimination argument for the incompatibility of externalism and self-knowledge is untenable because there is a conflict between its third premise and the standard incompatibilist justification for its second premise. In sections III and IV, I have defended this objection to Boghossian’s argument against various incompatibilist rejoinders. While doing this, I have examined in some detail the principles of relevance advanced by the chief proponents of the relevant-alternatives account of knowledge. Finally, in section V, three attempts to improve on the discrimination argument by reformulating it have been repudiated as unsatisfactory.